PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/62, C07K 3/18
A61K 37/43, C07K 13/00
C12P 21/02 //, C12N 15/16
A61K 37/24

A1

(11) International Publication Number: WO 92/10576

(43) International Publication Date: 25 June 1992 (25.06.92)

(21) International Application Number: PCT/US91/09152

(22) International Filing Date: 12 December 1991 (12.12.91)

(30) Priority data:
626,727 13 December 1990 (13.12.90) US

(60) Parent Application or Grant
(63) Related by Continuation
US 626,727 (CIP)
Filed on 13 December 1990 (13.12.90)

(71) Applicant (*for all designated States except US*): THE UPJOHN COMPANY [US/US]; 301 Henrietta Street, Kalamazoo, MI 49001 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*) : KUBIAK, Teresa, M. [PL/US]; 5844 East B Avenue, Richland, MI 49083 (US). SHARMA, Satish, K. [US/US]; 7649 Hampton Oaks, Portage, MI 49002 (US).

(74) Agent: DELUCA, Mark; Corporate Patents & Trademarks, The Upjohn Company, Kalamazoo, MI 49001 (US).

(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC (European patent), MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL (European patent), NO, PL, RO, SD, SE (European patent), SN (OAPI patent), SU+, TD (OAPI patent), TG (OAPI patent), US.

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: FUSION POLYPEPTIDES

(57) Abstract

A non-naturally-occurring fusion protein comprising an extension peptide portion covalently linked at its C-terminus to the N-terminus of a biologically active portion is disclosed. The extension peptide portion can be removed by DPP IV cleavage. A use of fusion proteins with DPP IV cleavable extension peptide portions in medecinal preparations is disclosed. A method of purifying desired proteins from a mixture containing a fusion protein is disclosed.

+ See back of page

FUSION POLYPEPTIDES

FIELD OF THE INVENTION

The present invention relates to non-naturally occurring fusion polypeptides containing N-terminal portions cleavable by dipeptidylpeptidase IV (DPP IV).

BACKGROUND OF THE INVENTION

The techniques of molecular biology, specifically recombinant DNA technology, allow for the production of relatively large quantities of desirable biologically active polypeptides. Furthermore, the genetic information encoding the polypeptides may be modified to produce relatively large quantities of modified polypeptides. Modifications made to the polypeptides are often used to improve their activity or facilitate their production and/or preparation. Accordingly, much effort has been made to determine what modifications are desirable in order to increase, enhance or otherwise alter the biological activity of desired polypeptides. In addition, there is a great deal of work being done to modify desired polypeptides to facilitate their production and purification.

Naturally produced polypeptides are often initially biosynthesized as larger precursors which are then trimmed by a series of proteolytic cleavages to produce the final products. Accordingly, several proteases exist which recognize and cleave specific amino acids and/or amino acid sequences. These proteases participate in a conversion of a precursor protein to the final polypeptide product.

Once such protease is dipeptidylpeptidase IV (DPP IV) (EC 3.4.14.5). DPP IV was first reported in Hopsu-Havu, V.K. and G.G. Glenner, Histo. Chemie 3:197-201 (1966) and has been shown to be present in many mammalian tissues. DPP IV is presently commercially available from Enzyme Systems Products (Dublin, California). DPP IV recognizes specific amino acid sequences on the N-terminus of proteins. Specifically, DPP IV will cleave a dipeptide from the N-terminus when the second amino acid from the N-terminus is proline (Pro), hydroxyproline (Hyp), alanine (Ala), serine (Ser), and threonine (Thr) and any amino acid is at the N-terminus residue position provided if proline or hydroxyproline is not the amino acid residue third from the N-terminus. DPP IV activity is more efficient when proline or alanine is the second amino acid from the N-terminus and is usually most efficient when that position is occupied by proline. The activity of DPP IV in the stepwise cleavage of "PRO" parts of precursors of naturally occurring peptides is widely reported.

Modern technology has made possible the high level production of biologically active proteins. Important polypeptides can be synthesized using peptide synthesizers or in host cells using recombinant DNA technology. Often, biologically active proteins are administered as drugs. Numerous examples exist in which active proteins are used as therapeutics, prophylactics or to enhance or repress traits. Since DPP IV and other proteases degrade proteins, these drugs

protein. When used as a prodrug, the non-naturally occuring protein is processed into a biologically active protein *in vivo* using DPP IV present in the target species. When used in a purification process, non-naturally occuring protein can be purified by using its specifically designed N-terminus as a ligand and then processed with DPP IV to remove the N-terminal extension and liberate a desired protein.

The present invention allows for the production of a desired protein as a non-naturally occuring protein that is later converted to the desired protein when exposed to DPP IV. Prodrugs are converted to drugs over a course of time using the patients' endogenous DPP IV, thereby achieving sustained presence of the active drug in a patient and reducing the frequency of administration. Pure desired proteins can be isolated using the present invention by producing and purifying non-naturally occuring proteins and then processing the non-naturally occuring proteins *in vitro* with DPP IV to produce the desired protein.

SUMMARY OF THE INVENTION

The present invention relates to a non-naturally occurring fusion protein comprising an extension peptide portion covalently linked at its C-terminus to the N-terminus of a core protein portion, said extension peptide portion being of the formula:

$$A\text{-}X\text{-}Y(X'\text{-}Y)_n$$

wherein

A is optional and when present is methionine;

n is 0-20;

X is selected from the group consisting of all naturally occurring amino acid residues;

X' is selected from the group consisting of all naturally occurring amino acid residues except proline and hydroxyproline;

Y is selected from the group consisting of proline, hydroxyproline, alanine, serine and threonine except when n is 0 then Y is selected from the group consisting of alanine, serine and threonine.

The present invention also relates to the use of such non-naturally occuring proteins in medicinal preparations and to a method of purifying desired proteins from a mixture containing such non-naturally occuring proteins and impurities comprising the steps of selectively contacting said non-naturally occuring protein with material which immobilizes said non-naturally occuring protein, removing said impurities, separating said non-naturally occuring proteins from said material, contacting said non-naturally occuring protein with DPP IV, and isolating said desired protein.

INFORMATION DISCLOSURE

U.S. Patent No. 4,569,794 issued February 11, 1986 to Smith et al relates to the process of purifying proteins and compounds useful in such processes. The invention describes

The affinity peptides disclosed contained at least two neighboring histidine residues. The IMAC purification means disclosed requires a special synthetic chemistry for making nitrilotriacetic acid (NTA) resins.

Tallon, M.A., et al., Biochem. 26:7767-7774 (1987) relate to synthesis of extended analogs of the tridecapeptide α-factor from *Saccharomyces cerevisiae*. The synthesized analogs are extended α-factors, which represent sequences of naturally occurring pro-α-factor coded for in the MFα1 structural gene.

Kriel, G. et al, Eur. J. Biochem. 111:49-58 (1980) describes the stepwise cleavage of the N-terminal portion of melittin precursor (Promelittin) by dipeptidylpeptidase IV. Promelittin is the main constituent of honeybee venom. In the amino acid sequence of the N-terminal portion of the precursor, every second residue is either proline or alanine. When promellitin is exposed to DPP IV isolated from pig kidney, the N-terminal region of the precursor is cleaved in a stepwise fashion producing the mature protein. Promelittin, unlike fusion proteins according to the present invention, is a naturally-occurring protein.

Julius, D., et al, Cell, Vol 32:839-852 (March 1983) relates to the role of membrane bound DPP IV in the processing of yeast α-factor from a larger precursor polypeptide. The yeast α-factor, unlike fusion proteins according to the present invention, is a naturally-occurring protein.

Mollay, C. et al, Eur. J. Biochem. 160:31-35 (1986) describes the isolation of DPP IV from the skin secretion of *Xenopus laevis*. The activity of DPP IV is discussed.

Mentlein, R., FEB, Vol. 234, No. 2, pp. 251-256 (July 1988) reviews proline residues in the maturation and degradation of peptide hormones and neuropeptides. It is reported that in mammals, proline specific proteases such as DPP IV are not involved in the biosynthesis of regulatory peptides but may play an important role in the degradation of them. Thus, it is concluded that while in vertebrates and lower vertebrates precursor proteins rely on DPP IV to convert precursors to mature forms, the processing of regulatory proteins in mammals generally uses DPP IV as a degradation protease.

Frohman, L. A. et al. J. Clin. Invest. 78:906-913 (1986) report that human growth hormone releasing factor (hGRF) and its analogs are rapidly degenerated *in vivo* in humans and *in vitro* by plasma DPP IV.

Frohman, L. A. et al. J. Clin. Invest. 83:1533-1540 (1989) report that human growth hormone releasing factor (hGRF) and its analogs are rapidly degenerated *in vivo* in humans and *in vitro* by plasma DPP IV.

Kubiak, T.M., et al, Drug Metabolism and Disposition, Vol. 17, No. 4, pp. 393-397 (1989) refer to the *in vitro* metabolic degradation of bovine growth hormone releasing factor (bGRF) analogs in bovine and porcine plasma and the correlation with plasma DPP IV activity.

Hochuli, E. et al., J. Chromat. 411:177-184 (1987) disclosed a nitrilotriacetic acid absorbent useful for metal chelate affinity chromatography. It is reported that the disclosed absorbent when charged with Ni^{2+} is useful in binding to peptides and proteins containing neighboring histidine residues.

Ljungquist, C. et al., Eur. J. Biochem. 186:563-569 (1989) disclose the use of the metal chelating peptide Ala-His-Gly-His-Arg-Pro in multiplicities of two, four and eight together with a column containing immobilized Zn^{2+} ions. According to Ljungquist use of this metal chelating peptide with zinc columns provides unexpectedly good purification of the fusion proteins.

DETAILED DESCRIPTION OF THE INVENTION

As used herein, the terms "non-naturally occurring fusion protein", "non-naturally occurring fusion polypeptide", "fusion polypeptides" and "fusion proteins" refer interchangeably to proteins and polypeptides which do not normally occur in nature and which comprise a core protein portion and an extension portion.

As used herein "core protein", "core protein portion" and "polypeptide portion" refers to the portion of a fusion polypeptide which is located at the C-terminus end of the molecule and which, absent the extension portion, would be a desired polypeptide and/or a biologically active protein including naturally occuring biologically active proteins and polypeptides and analogs and mutants thereof.

As used herein "N-terminal extension" refers to the first up to about 45 amino acids starting at the N-terminus and which are not part of the core protein.

As used herein "prodrugs" refers to fusion proteins wherein the biologically desired portion is a biologically active protein useful as a drug.

As used herein "biologically active protein" and "biologically active polypeptides" refer to interchangeable proteins and polypeptides which possess biological activity.

As used herein "desired protein" and "desirable protein" refer interchangeable to proteins and polypeptides which are sought in pure form.

As used herein "extension portion" refers to the portion of a fusion protein which is an N-terminal extension and which is not part of the biologically desired portion.

As used herein "DPP IV cleavable N-terminal extension portion" refers to the extension portion of a fusion protein which has an amino acid sequence that can be removed by the stepwise cleavage by DPP IV.

In the Sequence Listing Section, some amino acid residues have been designated Xaa in Seq ID. The following descriptions apply:

In Seq ID No. 3 Xaa^{29} represents C-terminally amidated Argininyl residue.

In Seq ID No. 4 Xaa^{29} represents C-terminally amidated Argininyl residue.

represents a covalent peptide bond; and, "core protein portion" represents any desired peptide which is liberated from the Extension portion by DPP IV processing.

The Extension portion of a fusion protein according to the present invention has an amino acid sequence according to the formula:

$$A\text{-}X\text{-}Y(X'\text{-}Y)_n$$

wherein A is optional, and when present is methionine;

n represents the number of sequentially linked X'-Y groups, that number representing from 0 to 20 of such groups, preferably 0 to 10 groups.

X is selected from the group consisting of any naturally occuring amino acid;

Y is selected from the group consisting of proline, alanine, serine, and threonine, except when $n = 0$, then Y is selected from the group consisting of alanine, serine, and threonine;

X' is selected from the group consisting of any naturally occuring amino acid except proline or hydroxyproline;

According to the formula, when $n = 1$, there are two Y residues. Further, it is possible to have up to twenty one Y residues and twenty X' residues in a single embodiment. Individual Y residues and X' residues respectively can be any residue of the group from which they are selected. That is, all of the individual Y residues do not have to be the same in a given embodiment. Similarly, in an embodiment with more than one X' residue, each individual X' residue present can be any amino acid residue except proline and hydroxyproline irrespective of what residue any other X' residue may be. Each individual Y and X' residue respectively must conform to the rules for that particular group and all that is necessary is that the various individual residues at the specific positions follow the rules as articulated above.

Fusion proteins in which (A) is present as methionine (Met) represent sequences useful for the production of biologically active proteins by recombinant DNA methods in *E. coli*. The Met sequence present in these precursors usually will be processed by the *E. coli* enzymatic system or some other means which can be performed by a person with ordinary skill in the art. Protein synthesis in *E. coli* is, under normal circumstances, initiated at the translation initiation codon AUG coding for Met. As a consequence, the newly synthesized polypeptides have a methionine residue as their N-terminal amino acid. *E. coli* possesses an enzymatic activity with the capacity to effectively remove N-terminal Met when the Met N-terminal residue is adjacent to an amino acid with a relatively small side chain like Gly, Ala or Ser as well as Pro. Highly specific removal of the N-terminal Met can be accomplished using cyanogen bromide mediated cleavage of Met. However, for that procedure to be successful, the N-terminal Met must be the only Met in the entire protein sequence; otherwise the cleavage will take place after each Met in the sequence. Accordingly, for fusion proteins containing internal Met sequences, the

will be used to process N-terminal extensions and, therefore, delay core protein degradation. That is, the extension portion of the fusion protein can act as a substrate for DPP IV and competitive inhibitor, delaying the DPP IV action on the core protein thereby temporarily protecting the core protein.

As used herein, "prodrug" means a fusion protein which contains a DPP IV cleavable N-terminal extension covalently linked to a core protein portion that is a biologically active polypeptide useful as a drug. Prodrugs according to the present invention can be administered as an individual proform or in combination with other compounds. The preferred embodiment is a well defined individual form of a prodrug. In either case, the proforms are processed by naturally occurring DPP IV normally found in the body.

The advantage of administering a prodrug in a medicinal preparation is that it delays activity and/or provides for extended presence of the biologically active protein. Prodrugs can remain active longer than unmodified molecules. Prodrugs can exist in a non-active state until such time elapses that a sufficient portion of the extension portion is degraded and the molecule becomes active. Prodrugs, therefore, can act as a time delayed drug delivery system. Furthermore, different N-terminal extensions are degraded at different rates, depending on their length and the specific residues present in their amino acid sequence. Combinations of different forms of prodrugs having a variety of N-terminal extensions can be provided which can provide a sustained, steady level of active drug in a patient over a course of time. Prodrugs, therefore, can act as a time delayed drug delivery system.

As described above, DPP IV cleaves off a dipeptide from the N-terminus of a polypeptide provided certain residues occupy certain positions. As used herein, "position one" refers to the amino acid residue position at the N-terminus. As used herein, "position two" refers to the amino acid residue position which is immediately adjacent to position one and which is second from the N-terminus. As used herein, "position three" refers to the amino acid residue position which is immediately adjacent to position two and which is third from the N-terminus. The cleavage which will remove the N-terminal dipeptide occurs between position two and position three provided amino acid three is not proline or hydroxyproline and amino acid two is one of five amino acids: proline (Pro), hydroxyproline (Hyp), alanine (Ala), serine (Ser), or threonine (Thr).

DPP IV cleaves the N-terminal residues at a different rate depending upon which of the four amino acid residues is present at position two. In most cases, DPP IV cleaves most efficiently when position two is occupied by Pro and it is next most efficient when position two is occupied by Ala. When position one is occupied by tyrosine, phenylalanine or histidine, DPP IV works at about the same rate when position two is occupied by Pro or Ala. DPP IV is next most efficient when position two is occupied by Ser. It is least efficient when Thr

according to the present invention. These polypeptides are meant only to serve as examples of embodiments and are not meant to limit the scope of the present invention.

Smaller fusion proteins according to the present invention can be synthesized, for example, by solid-phase methodology utilizing an Applied Biosystems 430A peptide synthesizer (Applied Biosystems, Foster City, California) as described in detail in PCT/US90/02923 and 07/368,231.

For larger molecules, production in host cells using recombinant DNA is preferred. There are several different methods available to one having ordinary skill in the art who wishes to use recombinant DNA technology to produce fusion proteins. Typically, genes encoding desired polypeptides are inserted in expression vectors which are then used to transform or transfect suitable host cells. The inserted gene is then expressed in the host cell and the desired polypeptide is produced. To produce the fusion polypeptides of the present invention in a like manner, an additional DNA sequence is included in the gene insert. Specifically, DNA encoding the N-terminal extension residues is operably linked to the 5′ end of the gene encoding the desired polypeptide. This additional genetic material must be placed downstream from the promoter of the expression vector so that it is under the control of the promoter. Additionally, it must be placed in proper reading frame with the gene so that the expressed protein product includes the N-terminal extension residues covalently linked to the desired polypeptide.

Therefore, to produce fusion proteins according to the present invention using recombinant DNA technology, oligonucleotides must be designed which encode the amino acid sequence of the desired N-terminal extension and these oligonucleotides must be operably inserted upstream of the 5′ end of the gene encoding the core protein portion, generating a chimeric gene. The techniques to make oligonucleotides and the techniques used to producing a chimeric gene are well known to those having ordinary skill in the art.

In addition to the utility of fusion proteins as prodrugs, the present invention relates to the purification and processing of biologically active recombinant polypeptides. The desired biologically active recombinant polypeptides are most preferably produced in a soluble form or secreted from the host. According to the present invention, the extension portion of the fusion protein can be recognized by purification means. The fusion protein is purified from the material present in the secretion media or extraction solution it is contained in and then processed to remove the extension portion from the core protein portion, thus producing purified desired protein. Accordingly, desired proteins most suited for processing as fusion proteins according to the present invention are those biologically active polypeptides which are not themselves substrates for DPP IV cleavage.

In accordance with the present invention, a gene sequence encoding for a desired

chelating a metal ion to the IDA-containing resin. The proteins bind to the metal ion(s) through functional groups of amino acid residues capable of donating electrons. Potential electron donating amino acid residues are cysteine, histidines, and tryptophan. Proteins interact with metal ions through one or more of these amino acids with electron donating side chains.

Smith et al. discloses in U.S. Patent No. 4,569,794, incorporated by reference herein, that certain amino acids residues are responsible for the binding of the protein to the immobilized metal ions. However, the bound protein can be eluted by lowering the pH or using competitive counter ligands such as imidazole if histidine side chains are involved in the binding. Histidine-containing di- or tripeptides in proteins have been used to show that IMAC is a selective purification technique. Accordingly, Smith et al. discloses using recombinant DNA techniques to produce a fusion protein comprising a metal chelating peptide covalently liked to a desired polypeptide. The metal chelating peptide is an extension portion that is effectively a handle to the desired polypeptide. This handle can be used in protein purification.

Use of IMAC technology with metal chelating peptides having alternating His residues is disclosed in U.S. Patent Application Serial No. 07/506,605, which is incorporated herein by reference. U.S. patent application Serial No. 07/506,605 discloses specific metal chelating peptides which provide unexpectedly superior results in the IMAC purification of a fusion protein when the metal chelating peptide comprises three to six alternating His residues. Following the teachings of U.S. patent application Serial No. 07/506,605 and U.S. Patent No. 4,569,794, it is possible to employ the commonly used IDA resin in IMAC for the purification of fusion proteins having a metal chelating peptide portion with at least three alternating histidine residues which are constituents of DPP IV-recognized sequences. Construction of fusion proteins and their use in an IMAC system is taught by U.S. Patent No. 4,569,794. Construction and use of a metal chelating peptide portion comprising alternating His residues is taught in U.S. Patent Application Serial No. 07/506,605. By providing a fusion protein with DPP IV recognizable residues between alternating His residues the present invention provides a fusion protein which can be purified using IMAC technology and subsequently processed with DPP IV to yield a desired polypeptide.

According to this embodiment of the present invention, the extension portion is a metal chelating peptide which can be represented by the formula:

$$\text{A-X-Y(X'-Y)}_n$$

and further, wherein A is optional, and when present is methionine;

n represents the number of sequentially linked X'-Y groups, that number representing from 0 to 20 of such groups, preferably 0 to 10 groups.

X is selected from the group consisting of any naturally occuring amino acid;

Y is selected from the group consisting of proline, alanine, serine, and threonine,

immobilized antibodies which recognize the antigenic portion of the fusion protein. The immobilized antibodies keep the protein in the column while the undesired components of the supernatant are eluted. The column conditions can then be changed to cause the antigen-antibody complex to dissociate.

According to the present invention, the highly antigenic N-terminal portion of the fusion protein is an extension portion which contains DPP IV recognizable residues. After collection as described in the Hopp patent, the fusion protein according to the present invention can be exposed to DPP IV, thereby removing the extension portion. One of ordinary skill in the art could practice the immunoaffinity purification system of Hopp with N-terminal extensions according to the present invention.

The embodiments and examples described herein serve to illustrate the nature of the present invention and are not meant to limit the scope of the invention. Contemplated equivalents include fusion proteins which have N-terminal extensions which can be processed by at least one other means such that removal of the extension is due to a combination of means. Contemplated equivalents also include fusion polypeptides comprising chemically modified amino acid residues.

EXAMPLES

Example 1 Synthetic Prodrugs which are Fusion Prodrugs Having Core Proteins that are DPP IV Substrates

Fusion polypeptides that can be synthesized and administered as prodrugs have a DPP IV degradable N-terminal extension covalently linked to the N-terminal of the biologically active polypeptide. The formula for these prodrugs can be expressed as the formula:

extension portion - core protein drug portion

wherein "extension portion" represents a DPP IV cleavable N-terminal extension; " - " represents a covalent peptide bond; and, "core protein portion" represents any desired peptide which is liberated from the extension portion by DPP IV processing. In this example the core protein of the fusion protein is a potential substrate for DPP IV following removal of the extension portion.

Synthetic prodrugs can be produced using peptide synthesis techniques well known in the art.

In one embodiment, the core protein portion is epidermal growth factor (EGF) and the extension portion is Gly-Pro-Phe-Ala:

$$Gly^{-4}\text{-}Pro^{-3}\text{-}Phe^{-2}\text{-}Ala^{-1}\text{-}[EGF].$$

In another embodiment, the core protein portion is glucagon and the extension portion is Ala-Pro-Phe-Ala:

$$Ala^{-4}\text{-}Pro^{-3}\text{-}Phe^{-2}\text{-}Ala^{-1}\text{-}[GLUCAGON].$$

A bGRF analog, Leu^{27}-bGRF $(1\text{-}29)NH_2$, its sequence shown as Seq ID 5, can be administered as a medicament comprising the core protein shown in Seq ID 5 and a variety of N-terminally extended prodrugs.

Several versions of prodrugs can be made by well known methods using Seq ID 5 as the core protein portion. Extension portions for these Seq ID 5-based prodrugs are Ile-Ala, Gly-Pro-Ile-Pro, Seq ID 6, Seq ID 7, Tyr-Ala, Gly-Pro-Tyr-Ala, Seq ID 8, Seq ID 9, Seq ID 10, Seq ID 11, Seq ID 12, Seq ID 13, Tyr-Ala-Tyr-Ala and Val-Ala.

Example 4 Sustained Presence of bGRF Analog $[Thr^2Ala^{15}Leu^{27}]$-bGRF $(1\text{-}29)NH_2$

A bGRF analog, $[Thr^2Ala^{15}Leu^{27}]$-bGRF $(1\text{-}29)NH_2$, its sequences shown as Seq ID 14, can be administered as a medicament comprising the core protein portion shown in Seq ID 14 and a variety of N-terminally extended prodrugs. Three versions of the prodrug were made having extension portions of Tyr-Thr, Tyr-Ser, and Tyr-Thr-Tyr-Thr, respectively.

Example 5 Fusion Proteins which Contain HIV RNase H and N-Terminal Extensions

A strategy to purify chimeric proteins from recombinant E. coli is described based on metal chelating peptide domains containing alternate histidines, with affinity for an immobilized metal ion. Vectors are constructed to direct the synthesis of fusion proteins using HIV RNase H as the core protein. As shown below, these fusion proteins are designed to possess alternating histidines for purification by immobilized metal ion affinity chromatography (IMAC) and alternating prolines or alternating alanines for DPP IV cleavage to remove the metal chelating peptide (mcp).

The preferred DPP IV cleavable N-terminal extensions according to the present invention are outlined as follows:

Fusion protein HIVRH/mcp #1 comprises an N-terminal extension of Seq ID 15 linked to HIV RNase H:

Met^{-11}-Pro^{-10}-Ala^{-9}-His^{-8}-Pro^{-7}-His^{-6}-Pro^{-5}-His^{-4}-Pro^{-3}-His^{-2}-Ala^{-1}-[HIV RNase H]

Fusion protein HIVRH/mcp #2 comprises an N-terminal extension of Seq ID 16 linked to HIV RNase H:

Met^{-11}-Ala^{-10}-Pro^{-9}-His^{-8}-Ala^{-7}-His^{-6}-Ala^{-5}-His^{-4}-Ala^{-3}-His^{-2}-Ala^{-1}-[HIV RNase H]

Fusion protein HIVRH/mcp #3 comprises an N-terminal extension of Seq ID 17 linked to HIV RNase H:

Met^{-11}-Gly^{-10}-Pro^{-9}-His^{-8}-Pro^{-7}-His^{-6}-Pro^{-5}-His^{-4}-Pro^{-3}-His^{-2}-Ala^{-1}-[HIV RNase H]

These fusion proteins are cloned and expressed in *E. coli*, and are purified using DEAE chromatography and RP-HPLC. N-terminal sequencing is used to characterize the fusion

Pharmacia is washed thoroughly with Milli-Q water on a scintered glass filter. The gel is then resuspended in water to form a slurry. The slurry is poured carefully into a glass column (Pharmacia) to a volume of 6 mls (1 x 7 cm). After the gel has settled, the column is washed with 5 volumes of 50 mM EDTA (ethylenediaminetetraacetic acid) pH 8.0. Following this, the column is washed with 5 volumes of 0.2 N NaOH and 5 volumes of Milli-Q water. The column then is charged with 5 volumes of 50 mM $NiSo_4$ (or $ZnCl_2$ or $CuSO_4$). Finally, the column is washed with 5 bed volumes of equilibration buffer. The equilibration buffer is made up of 20 mM Tris pH 8.0, containing 500 mM NaCl, 1 mM PMSF, 1 mM benzamidine, 10 mg/L leupeptin, and 10 mg/L aprotinin.

The column has been equilibrated with at least 5 volumes of equilibration buffer. 5-10 mls of crude recombinant *E. coli* extract are applied to the column by gravity. After all the crude material has entered the column, the column is washed with 10 column volumes of equilibration buffer containing 1.0 M NaCl, instead of 500 mM NaCl, pH 8.0.

The column is then eluted with increasing concentrations of imidazole in the equilibration buffer at pH 8.0. For the earlier experiments, a large number of elutions are performed for each experiment to determine the concentration at which the chimeric eluted. Later this elution is simplified and usually just three imidazole concentrations are used: 35 mM, 100 mM, and 300 mM imidazole in the equilibration buffer, pH 8.0. Ten bed volumes of each imidazole buffer are used. Between elutions, the column is washed with 10 volumes of equilibration buffer. Finally, the column is stripped with 5 bed volumes of 50 mM EDTA pH 8.0 to determine if any protein is still bound to the column. The flow rates for the columns are 1.0 ml/min. 5 ml fractions are collected. The columns are run at room temperature.

Commercially available Pierce protein assay kits are used to determine the protein content of the samples.

HIV RNase H activity is determined by the method described in Becerra, S. P. et al, FEBS 270(1,2):76-80 (September 1990), incorporated herein by reference.

Conversion of the N-terminal extended fusion proteins to mature proteins

Commercially available DPP IV purified from human placenta (Enzyme Systems Products, Dublin, Ca.) with a specific activity of 5200 mU per mg protein is used. One U is equivalent to hydrolysis of 1 umole of a synthetic substrate, Ala-Pro-7-amino-4-2 trifluoromethyl coumarin in one minute at pH 7.8. Enzymatic conversion is carried out by incubation of the fusion protein (about 1-100 mg) at a concentration of 1-10 mg/ml with DPP IV at 25 degrees C for 30 minutes at an enzyme to substrate ratio of 1:100 (w/w). The desired polypeptide is recovered from the uncleaved fusion protein by IMAC. The authenticity is confirmed by N-terminal sequence analysis.

Example 6. Processing of bGRF Analog prodrugs in bovine plasma *in vitro*.

prodrugs having either 4 or 2 amino acids in the extension (peptides Seq ID 19 and Seq ID 18, respectively) despite the significant difference in the *in vitro* half-life of the core peptide generated from these two bGRF prodrugs *in vitro* as indicated in Table 1. The *in vivo* growth hormone release was rapid and the same for the core peptide Seq ID 5 as well as for Seq ID 19 and 18, with no difference in the time of the GH peak following the challenge with bGRF analogs.

Our interpretation of these results is that most likely the rapid GH release from prodrugs *in vivo* is due to the overall high tissue and organ DPP-IV levels which are ca. 100 fold higher than the plasma DPP-IV concentration. This will explain the difference in the rate of prodrugs processing *in vivo* as compared with the cleavages observed in the in vitro experiments summarized in Table 1. It is also feasible that the half-life of the core peptide generated from its prodrug precursors was extended in vivo but not sufficiently to show an altered (extended) growth hormone release. It is known that GH release *in vivo* is modulated by the stimulatory effect of bGRF and inhibitory action of somatostatin (somatotropin release inhibitory factor, SRIF). In the meal-fed steer model of Moseley, et al. (J. Endocr. 17, 253-259, 1988), animals are injected iv with GRF two hours prior to feeding for the reason that the pituitary is more responsive to a GRF challenge before feeding versus following feeding. Factors associated with feeding such as release of gut/pancreatic SRIF may interfere with the ability of the pituitary to release GH. In other words, no GH will be released from the pituitary even in the presence of GRF during the SRIF overtone. Normally, in the unchallenged meal-fed steer model, serum GH concentration declines to basal levels for 3 to 6 hrs after feeding (so called trough period) with another exogenous episodic GH pulse at 5 to 8 hrs following feeding. In response to GRF injection 2 hrs before feeding, the GH response is rapidly occurring within 5-20 min. and GH level remains elevated for 120 to 240 min. before returning to baseline. In the case of GRF prodrugs tested so far, only the first exogenous GH peak was elevated, the second one did not show any treatment effect. It is possible that the half-life of the core GRF generated from prodrugs in our experiments was not extended long enough to allow for the core peptide to be present in the circulation in sufficient concentrations to affect the second exogenous burst of exogenous GH, usually 4-6 hours after the first one.

Taken together, our results support the general prodrug concept disclosed here because: (i) bGRF prodrugs having DPP-IV-cleavable N-terminal extensions were processed successfully to produce the core peptide(s) in bovine plasma in vitro via DPP-IV mediated cleavages; (ii) the in vitro half-life of the core peptide generated from the prodrugs was significantly longer and was a function of the N-terminal extension length in the prodrug; (iii) the fact that the bGRF prodrugs with very low inherent potency were as effective as the core peptide in the release of GH in vivo indicates that most likely the core peptide was generated in vivo as anticipated.

Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.01 (4); Thr 0.96 (1); Ser 1.80 (2); Glu 2.02 (2); Pro 0.97 (1); Gly 1.98 (2); Ala 3.91 (4); Val 0.99 (1), Ile 1.89 (2), Leu 5.08 (5); Tyr 3.05 (3); Phe 0.98 (1); Lys 2.03 (2); Arg 3.06 (3).

Example 12 Preparation of Tyr^{-6}-Ala^{-5}-Gly^{-4}-Pro^{-3}-Tyr^{2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 30 which comprises Seq ID 8 as the extension portion and which has the formula:

#7 Tyr-Ala-Gly-Pro-Tyr-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.07 (4); Thr 0.96 (1); Ser 1.79 (2); Glu 2.02 (2); Pro 0.99 (1); Gly 1.95 (2); Ala 4.80 (5); Val 0.96 (1), Ile 1.87 (2), Leu 5.09 (5); Tyr 4.11 (4); Phe 0.97 (1); Lys 2.06 (2); Arg 3.08 (3).

Example 13 Preparation of Lys^{-8}-Pro^{-7}-Tyr^{-6}-Ala^{-5}-Gly^{-4}-Pro^{-3}-Tyr^{-2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 31 which comprises Seq ID 9 as the extension portion and which has the formula:

#8 Lys-Pro-Tyr-Ala-Gly-Pro-Tyr-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.06 (4); Thr 0.95 (1); Ser 1.78 (2); Glu 2.01 (2); Pro 1.95 (2); Gly 1.96 (2); Ala 4.81 (5); Val 0.95 (1), Ile 1.87 (2), Leu 5.09 (5); Tyr 4.12 (4); Phe 0.96 (1); Lys 3.08 (3); Arg 3.10 (3).

Example 14 Preparation of Phe^{-10}-Ala^{-9}-Lys^{-8}-Pro^{-7}-Tyr^{-6}-Ala^{-5}-Gly^{-4}-Pro^{-3}-Tyr^{-2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 32 which comprises Seq ID 10 as the extension portion and which has the formula:

#9 Phe-Ala-Lys-Pro-Tyr-Ala-Gly-Pro-Tyr-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino

(2), Leu 5.09 (5); Tyr 4.10 (4); Phe 1.98 (2); Lys 3.03 (3); Arg 4.09 (4).

Example 18 Preparation of Val^{-2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 36 having the formula: #14 Val-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 3.98 (4); Thr 0.89 (1); Ser 1.76 (2); Glu 2.02 (2); Gly 1.05 (1); Ala 3.87 (4); Val 1.85 (2), Ile 1.77 (2), Leu 5.17 (5); Tyr 2.04 (2); Phe 0.97 (1); Lys 2.07 (2); Arg 3.06 (3).

Example 19 Preparation of Tyr^{-2}-Thr^{-1} {[Ala^{15} Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 37 having the formula: #15 Tyr-Thr-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Ala-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.06 (4); Thr 1.86 (2); Ser 1.77 (2); Glu 2.07 (2), Ala 3.98 (4); Val 1.08 (1), Ile 1.89 (2), Leu 5.14 (5); Tyr 2.94 (3); Phe 0.96 (1); Lys 1.99 (2); Arg 3.04 (3).

Example 20 Preparation of Tyr^{-2}-Thr^{-1} {[Ile^2 Ala^{15} Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 38 having the formula: #16 Tyr-Thr-Tyr-Ile-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Ala-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.07 (4); Thr 1.87 (2); Ser 1.75 (2); Glu 2.07 (2), Ala 2.94 (3); Val 1.09 (1), Ile 2.87 (3), Leu 5.12 (5); Tyr 2.92 (3); Phe 0.96 (1); Lys 2.00 (2); Arg 3.05 (3).

Example 21 Preparation of Tyr^{-2}-Thr^{-1} {[Thr^2 Ala^{15} Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 39 having the formula: #17 Tyr-Thr-Tyr-Thr-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Ala-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.05 (4); Thr 2.68 (3); Ser 1.77 (2); Glu 2.07 (2), Ala 2.90 (3); Val 1.08 (1),

parantheses: Asp 3.97 (4); Thr 0.90 (1); Ser 1.74 (2); Glu 1.98 (2); Gly 1.04 (1); Ala 4.85 (5); Val 0.91 (1), Ile 1.77 (2), Leu 5.13 (5); Tyr 4.14 (4); Phe 0.99 (1); Lys 2.07 (2); Arg 3.05 (3).

Table 2. Serum GH Response to IV Injections of Various Doses of [Leu^{27}]-bGRF(1-29)NH_2 (Seq ID 5) and Ile^{-2}-Pro^{-1}{[Leu^{27}]-bGRF(1-29)NH_2} (Seq ID 18) in Meal-Fed Holstein Steers.*

Treatment	Dose nmol/kg	Number of Animals Responding	Peak Height (ng/ml)		Time to Peak (min)		Area 0-8 h (Unit)	
			A[a]	B[a]	A[a]	B[a]	A[a]	B[a]
Saline	0	0^{b}	32.4^{b}	32.4^{b}	89^{b}	89^{b}	4.3^{b}	4.3^{b}
Seq ID 18	0.02	$8/10^{c}$	$71.2^{b,c}$	$76.9^{b,c}$	23^{c}	18^{c}	$4.6^{b,c}$	$5.0^{b,c}$
Seq ID 5	0.20	$9/10^{c}$	119.8^{c}	130.9^{c}	23^{c}	14^{c}	6.8^{d}	7.0^{d}
Seq ID 18	0.20	$10/10^{c}$	$101.4^{c,d}$	101.4^{c}	26^{c}	26^{c}	$6.3^{c,d}$	$6.3^{c,d}$
Seq ID 18	20.0	$10/10^{c}$	137.8^{d}	137.8^{c}	23^{c}	23^{c}	10.1^{e}	10.1^{e}
SEM		.04	9.1	9.4	8	8	.3	.3
p Value		.0001	.007	.007	.04	.03	.0001	.0001

* Animals were injected IV with peptides at the doses indicated 2 hrs before feeding and procedures were as described by Moseley et al. J. Endocrinology 117:253-259 (1988).

[a] Analysis A includes all steers and Analysis B includes only steers responding to GRF injection and control steers.

[b,c,d,e] Values with different superscripts in a column are significantly different ($P<.05$).

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Kubiak, Teresa M.
Sharma, Satish K.

(ii) TITLE OF INVENTION: Fusion Polypeptides

(iii) NUMBER OF SEQUENCES: 42

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Upjohn Company - Corp. Patents & Trademarks
(B) STREET: 301 Henrietta Street
(C) CITY: Kalamazoo
(D) STATE: Michigan
(E) COUNTRY: USA
(F) ZIP: 49001

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: diskette (3M 3.5, DS double side 1.0 MB)
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: WordPerfect 5.1

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US07/626,727
(B) FILING DATE: 13/12/90

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US07/614,170
(B) FILING DATE: 14/11/90

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US90/02923
(B) FILING DATE: 30/05/90

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US07/368,231
(B) FILING DATE: 16/06/89

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US07/506,605
(B) FILING DATE: 09/04/90

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: DeLuca, Mark
(B) REGISTRATION NUMBER: 33229
(C) REFERENCE/DOCKET NUMBER: 4595

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 616 385 5210
(B) TELEFAX: 616 385 6897

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
20 25

(5) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 29

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa29

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Tyr Ile Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Ala Gln
1 5 10 15

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
20 25

(6) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 29

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa29

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Gly Gln
1 5 10 15

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
20 25

(7) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 6

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Tyr Ala Gly Pro Ile Pro

(12) INFORMATION FOR SEQ ID NO:11:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 12
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala
1 5 10

(13) INFORMATION FOR SEQ ID NO:12:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 14
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Val Pro Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala
1 5 10

(14) INFORMATION FOR SEQ ID NO:13:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 16
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Arg Pro Val Pro Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala
1 5 10 15

(15) INFORMATION FOR SEQ ID NO:14:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 29
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa29
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Tyr Thr Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Ala Gln
1 5 10 15

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa

Ile Pro Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu
1 5 10 15

Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
20 25 30

(20) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 33

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa33

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Tyr Ala Tyr Ala Tyr Ala Asp Ala Ile Phe Thr Ser Ser Tyr Arg Lys
1 5 10 15

Val Leu Ala Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Ser
20 25 30

Xaa

(21) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 39

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa39

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala Tyr Ala Asp Ala Ile Phe
1 5 10 15

Thr Asn Ser Tyr Arg Lys Val Leu Ala Gln Leu Ser Ala Arg Lys Leu
20 25 30

Leu Gln Asp Ile Leu Asn Xaa
35

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa27

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Gly Gln Leu Ser
 1               5                  10                  15

Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
            20                  25
```

(26) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 31

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa31

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
Tyr Ala Tyr Ala Asp Ala Ile Phe Thr Ser Ser Tyr Arg Lys Val Leu
 1               5                  10                  15

Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
            20                  25                  30
```

(27) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 33

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa33

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
Gly Pro Ile Pro Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys
 1               5                  10                  15

Val Leu Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn
            20                  25                  30
```

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa33
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Gly Pro Tyr Ala Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys
1 5 10 15

Val Leu Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn
20 25 30

Xaa

(31) INFORMATION FOR SEQ ID NO:30:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 35
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa35
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Tyr Ala Gly Pro Tyr Ala Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr
1 5 10 15

Arg Lys Val Leu Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile
20 25 30

Leu Asn Xaa
35

(32) INFORMATION FOR SEQ ID NO:31:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 37
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa37
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Lys Pro Tyr Ala Gly Pro Tyr Ala Tyr Ala Asp Ala Ile Phe Thr Asn

35 40

(35) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 43

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa43

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
Val Pro Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala Tyr Ala
 1               5                  10                  15

Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Gly Gln Leu Ser
            20                  25                  30

Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
        35                  40
```

(36) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 45

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
Arg Pro Val Pro Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala
 1               5                  10                  15

Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Gly Gln
            20                  25                  30

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
        35                  40                  45
```

(37) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 31

(40) INFORMATION FOR SEQ ID NO:39:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 31
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa31
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

```
Tyr Thr Tyr Thr Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu
 1               5                  10                  15

Ala Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
            20                  25                  30
```

(41) INFORMATION FOR SEQ ID NO:40:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 31
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa31
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

```
Tyr Ser Tyr Thr Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu
 1               5                  10                  15

Ala Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
            20                  25                  30
```

(42) INFORMATION FOR SEQ ID NO:41:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 33
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa33
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

```
Tyr Thr Tyr Thr Tyr Thr Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys
```

CLAIMS

1. A non-naturally-occurring fusion protein comprising an extension peptide portion covalently linked at its C-terminus to the N-terminus of a core protein portion, said extension peptide portion being of the formula:

$A\text{-}X\text{-}Y(X'\text{-}Y)_n$

wherein

A is optional and when present is methionine;

n is 0-20;

X is selected from the group consisting of all naturally occurring amino acid residues;

X' is selected from the group consisting of all naturally occurring amino acid residues except proline and hydroxyproline;

Y is selected from the group consisting of proline, hydoxyproline, alanine, serine and threonine except when n is zero and A is absent then Y is selected from the group consisting of alanine, serine and threonine.

2. A non-naturally-occurring fusion protein according to claim 1 wherein A is present and X is selected from the group consisting of Pro, Gly, Ala and Ser.

3. A non-naturally-occurring fusion protein according to Claim 1 wherein n is 0-10.

4. A non-naturally-occurring fusion protein according to claim 1 wherein said biologically active polypeptide is selected from the group consisting of: bGRF analogs, EGF; IGF-2, glucagon; corticotropin releasing factor; dynorfin, somatostatin-14; endothelin; transforming growth factor α; Vasoactive Intestinal Peptide; human β-casomorphin; Gastric Inhibitory Peptide; Gastric Releasing Peptide; human Peptide HI; human Peptide YY; glucagon-like peptide-1 fragment 7-37; glucagon-like peptide-2; substance P; Neuropeptide Y; human Pancreatic Polypeptide; insulin-like growth factor-1; human growth hormone; bovine growth hormone; porcine growth hormone; prolactin; human growth hormone releasing factor; bovine growth hormone releasing factor; porcine growth hormone releasing factor; ovine growth hormone releasing factor; interleukin -1β; and interleukin-2.

5. A non-naturally-occurring fusion protein according to claim 1 wherein said extension peptide portion is selected from the group consisting of Gly-Pro-Ile-Pro, Seq ID 6, Seq ID 7, Tyr-Ala, Gly-Pro-Tyr-Ala, Seq ID 8, Seq ID 9, Seq ID 10, Seq ID 11, Seq ID 12, Seq ID 13, Tyr-Ala-Tyr-Ala, Val-Ala, Seq ID 15, Seq ID 16, Seq ID 17, Seq ID 22 and Seq ID 23.

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 91/09152

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12N15/62; C07K3/18; A61K37/43; C07K13/00
C12P21/02; // C12N15/16; A61K37/24

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12N ; C07K ; A61K ; C12P

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
X,P	WO,A,9 015 821 (THE UPJOHN COMPANY) 27 December 1990 cited in the application see the whole document, and especially : page 6, line 36 - page 7, line 16, and page 11, lines 4-16	1,3-6, 10,12-13
X	EP,A,0 184 355 (ELI LILLY AND COMPANY) 11 June 1986	1,3-4
A	see page 3, line 16 - page 6, line 22 see page 18; table I & US,A,4 569 794 (ELI LILLY AND COMPANY) 11 February 1986 cited in the application	2,7-9, 13-14,16

-/--

° Special categories of cited documents : [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
30 MARCH 1992	14. 04. 92
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	ANDRES S.M.

2

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. US 9109152 SA 54941

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on 30/03/92
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9015821	27-12-90	AU-A- 5674390	08-01-91
EP-A-0184355	11-06-86	US-A- 4569794	11-02-86
		AU-B- 588840	28-09-89
		AU-A- 5024085	12-06-86
		CA-A- 1252948	18-04-89
		JP-A- 61148197	05-07-86
US-A-4569794	11-02-86	AU-B- 588840	28-09-89
		AU-A- 5024085	12-06-86
		CA-A- 1252948	18-04-89
		EP-A,B 0184355	11-06-86
		JP-A- 61148197	05-07-86
EP-A-0282042	14-09-88	AU-B- 609783	09-05-91
		AU-A- 1270988	15-09-88
		JP-A- 63251095	18-10-88
		ZA-A- 8801534	12-09-88
EP-A-0150126	31-07-85	US-A- 4703004	27-10-87
		DE-A- 3584260	07-11-91
		JP-A- 60222429	07-11-85
		US-A- 4782137	01-11-88
US-A-4782137	01-11-88	US-A- 4703004	27-10-87
		DE-A- 3584260	07-11-91
		EP-A,B 0150126	31-07-85
		JP-A- 60222429	07-11-85
WO-A-9115589	17-10-91	AU-A- 7685791	30-10-91

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82